

图书馆数字资源采集机制初探

黄建年 陶茂芹

(南京财经大学 南京 210046)

摘 要: 图书馆该如何采集数字资源,以形成完善的文献资源保障体系是图书馆数字资源建设的一个重要内容。从策略、原则、内容等方面详细描述了图书馆的数字资源采集机制。

关键词: 数字资源 采集 数字图书馆 资源建设

LIBRARY'S COLLECTING SYSTEM OF DIGITAL RESOURCES

Huang Jiannian Tao Maoqin

(Nanjing University of Finance & Economics Nanjiang 210046)

Abstract: One of the important content of the construction of digital resources in library is collecting of the digital resources to form the perfect literature resources guarantee system. The digital resource collecting system of library has been discussed from the points of the strategy, principle, content and so on.

Key words: digital resource collection digital library resource construction

随着科学技术的飞速发展,文献资源总量迅速增加,数字资源也日益增多。比如在国家 973 和 985 项目支持下的中国网页历史信息存储与展示系统——“中国 Web 信息博物馆”目前已经维护有近四亿中文为主的网页,并以平均每月一千万网页的速度扩大规模,发展速度实在惊人。馆藏文献的数字化也已经成为我国图书馆发展的一个重要趋势,大量的数字文献纷纷产生。“中国实验型数字图书馆”科研项目启动后,每年有 3 000 万页数字化文献将面世。不仅如此,各学科文献资源的交叉、重复、分散、老化加速的现象也日趋严重。面对如此众多的、交叉、重复、分散的数字化资源,要找到针对性很强、价值很高的资源,的确是一件不容易的事。为此,各图书馆应该大力加强数字资源的采集、整理、加工,以形成完整的、全方位的文献资源保障体系。因篇幅所限,此文仅就图书馆的数字资源采集机制作一探讨,希望能为图书馆数字资源建设提供一些思路。

1 数字信息资源采集涵义

1.1 数字信息资源概述

信息资源是指包括信息基础设施、信息数据、信息人才、信息技术、文献资源在内的信息生产、流通、分配、使用全过程中所有信息要素的总和。因为涉及的范围太广泛,我们不可能面面俱到。为了叙述的方便,此处我们说的资源主要是指文献信息资源。

图书馆数字信息资源一般是指以电子数据的形式将文字、图像、声音、动画等多种形式的信息存储在光磁等非印刷纸质的载体中,并通过网络通信、计算机或终端等方式再现出来的信息资源。就目前情况看,图书馆的数字信息资源可分为两大部分:一是本馆的现实数字化馆藏,比如光盘等电子出版物等;二是虚拟馆藏,即是网络上的信息资源。在图书馆计算机网络发展的不同阶段,其馆藏数字信息资源的构成是不尽相同的。在微机局域网阶段,馆藏书目信息数据库、读者数据库和流通数据库是图书馆主要的数字化资源,辅以一定数量的单机的磁盘、光盘等电子出版物。而到了现阶段,除了自建的各种数据库外,还包括了网络版的各种电子出版物等,同时还应包括经

过系统化处理的网络信息资源。如何处理好三种资源之间的关系,工作重点在何处,利用何种采集策略,这些问题都应该引起我们的注意和思考。

1.2 图书馆数字信息资源采集

根据本馆特点以及学科重点,按照一定的方法和原则,全面、完整、系统地采购、搜集各种数字资源,以形成数字资源保障体系,并最终形成完善的文献资源保障的过程,我们可以称为数字资源采集。

2 数字资源采集的原则

2.1 目的性原则

数字信息采集根据总目标,确定数字信息收藏体系的整体结构,对数字信息的学科范围、类型、结构层次等的总体设计或规划设想。图书馆信息资源的采集,应根据本单位的性质、方向任务和服务对象进行采集工作。紧密结合图书馆的方针、任务及学科规划,收集数字资源既要全面系统,又要有针对性,做到有的放矢。

2.2 系统性原则

数字信息资源收藏的系统性和完整性是指知识门类上其系统性和完整性。从纵系统而言,凡属服务的学科的重要资源,要尽量收集得系统完整,从而保持学科的连贯性;从横系统而言,要研究和照顾这些学科与其他学科领域的边缘交错,相互渗透的情况,要反映出这些学科间相互关联、互为条件的客观规律,从而保证资源的实用价值。

2.3 连续性、稳定性原则

数字信息资源收藏是在明确目的性原则的指导下持续不断地补充、长期的积累。只有这样才能反映这些学科资源的历史、学科的发展状况、特点及规律,从而保证所收集资源具有较高的使用价值。

2.4 信息及时准确原则

当代科学技术的发展一日千里,同时数字信息资源具有传播迅速的特点,为此,数字信息资源的收集机构更应力争在最短的时间内向用户提供最新、又为他们所急需的资源。这要求收集人员有较高的工作效率和强烈的时间观念。数字资源采集还应该有一定的预见性。收集人员要密切注意本学科及相关学科的发展水平与动向,注意新兴学科、交叉学科和边缘学科的产生与发展,从而增强服务的主动性。

2.5 开放性原则

随着科学技术的高速发展,知识量急剧增加,资源呈“爆炸”趋势,另一方面用户对信息资源的需求具有广泛性和多层次性,信息资源的建设已经走向相互

依存、统一规划、分工协作和资源共享的道路。数字信息资源的收集,从自己的特点出发,根据整体布局,各有侧重,避免重复,共享资源,从而建设成一个极其丰富的开放的信息资源系统。

3 数字资源采集策略

3.1 明确目标

图书馆信息资源的采集,应根据本单位的性质、方向任务和服务对象进行采集工作。紧密结合本馆的方针、任务及学科规划,收集数字资源既要全面系统,又要有针对性。我院是经济类院校,应在经济类专业方面逐步形成一定的特色数字资源收藏体系,并成为未来全国虚拟图书馆的一个重要组成部分。

3.2 确定任务

1) 调查本单位的方向任务及用户需求;定期了解服务对象的科研任务、计划、规划、目前工作重点、急需解决的关键性问题及其应用情况。通过各种方式,广泛征求用户对于资源的需要及其利用情况的意见。根据数据流量和咨询工作记录等,经常对资源情况进行分析研究,然后将结论性意见纳入数字资源的补充计划。

2) 调查数字资源的出版发行情况和搜集渠道。充分掌握国内外那些与本单位收藏重点有关的各出版发行机构的基本情况,以便及时补充最新的数字资源。

3) 网络资源的搜集工作。在数量庞大的网络信息资源中,实用性的只占很小的一部分。这就需要在搜集过程中,利用各种搜索工具,来采集那些实用性强的信息,从而使本单位数字信息资源建设得到完善。

4) 调查本单位非数字资源入藏情况。在搜集过程中,还必须经常地、有计划、有重点地对非数字资源进行分析研究,以保证资源收藏的系统性和完整性,并能够形成数字资源与非数字资源相互配合、相互补充的文献保障体系。

3.3 数字资源采集的程序

数字资源收集工作的程序即指搜集资源的具体工作步骤,其制定以工作规律为依据,不能随意变动。程序大体可分为以下几个步骤:

1) 搜集资源目录,通过各种途径,采取多种方式,搜集和积累各类资源目录是首要环节,可通过发行部门、数字信息资源广告以及网络上各种搜索工具等,以作为选择的线索。

2) 圈选资源,根据资源收集的既定原则、范围和

规划,来选订各类资源。

3) 检查,首先核查馆藏资源中是否有已有该资源,然后核查在声音质量、图象大小、图象质量等方面的效果。

4) 送审,将已圈好的资源目录送有关人员和上级业务部门审查,特别对高价数字资源,更应严格审批手续。送审是最后的决策过程。

5) 办理预订手续,向有关部门办理相关手续,比如订购、汇款、签订相关使用协议、向产品提供商提供本馆、本校 IP 地址等。

6) 验收以及登录。检查获得的资源是否为本馆采集的资源,如果是,数量品种、使用权限是否相符,核对无误后,进行相应的财产登记工作。

4 数字资源采集机制具体内涵

4.1 数字资源类型

1) 按数字资源的形式分为:图书、期刊、会议资源、科技报告、专利文献、标准文献、政府出版物、学位论文、产品资料、科技档案以及内部资料。

2) 按数字资源的格式分为:文本、图片、录音、录像、电影、通用软件、课件、习题、试题、案例等。

3) 按数字资源的层次分为:基本元素、基本知识点和组合知识点。基本元素即知识原子,是形成知识点的文本、图形、图像、声音等基本单元。基本知识点则是由基本元素构成的论题。而组合知识点则由基本知识点构成的模块。

4) 按数字资源的结构分为:一次文献、二次文献、三次文献。一次文献是以作者本人的科研、生产和工作成果为依据的原始文献,它具有创造性、原始性、分散性以及使用价值较高等特点,比如各单位自己发布的产品信息、服务项目等;二次文献是人们为了积累知识,将大量分散的、无组织、无系统的一次文献采用一定的科学方法进行加工、整理、提炼和浓缩汇集而成的系统的、完整的、便于检索和提取的文献,它具有汇集性、工具性、系统性等特点;三次文献是指动态综述、专题述评、进展报告、数据手册、书目之书目等,它具有综合性、针对性、价值性的特点。

4.2 数字信息源

4.2.1 传统数字信息源

1993 年世界电子出版物的总数为 8 000 多种,1994 年超过了 10 000 种,1995 年超过了 15 000 种,形成了一个规模巨大的电子出版业,在市场发展方面有关专家预测世界电子出版物的市场销售将以每年 30%~40% 的速度增加。光盘出版物也发展很快。

据英国伦敦 TFL 公司出版的《光盘指南》统计,1989 年全世界光盘出版物为 240 种,1990 年则发展为 817 种,1991 年为 1 400 种,1992 年为 2 212 种,1993 年为 3 500 种,1994 年达 8 200 种,1995 年为 9 500 种左右。所以说,搜集这些传统的电子出版物很有必要。

4.2.2 网络数字信息源

1) 大型学术信息系统,例如:中国知识基础设施工程(NIKI) (<http://cnki.net/indexie4.htm>);中国高等教育文献保障体系(CALIS) (<http://www.calis.edu.cn/>);万方数据信息资源系统(<http://chinainfo.gov.cn/indexs.htm>)等资源。

2) 网上出版物、出版社及网上书店。在因特网上的电子报纸和电子杂志数量增长迅速,目前已达数千种,各国报纸和杂志的出版商也都纷纷向电子化方向转移。例如:情报学报(<http://www.chinainfo.gov.cn/periodical/qbxb/in2dex.html>);中国信息导报(<http://www.chinainfo.gov.cn/periodical/zgxxdb/index.htm>);出版社主要有国家图书馆出版社(<http://nlc.gov.cn/disk4/cbs/indx.html>)等。

3) 联机数据库,比如世界知名的 DIALOG 系统(<http://www.dialog.com/>);中国学位论文数据库(<http://db.sti.ac.dn/demo/search-cddb.html>)等;中国期刊网(<http://www.chinajournal.net.cn>)。

4) 各专题网址。例如:<http://www.carl.org/uncover.html> 提供 1.7 万种英文期刊的免费检索和专题文章收费拷贝服务;<http://www.isleuth.com> 可进行专题数据库搜索;<http://www.elsevier.nl> 对新出版和即将出版的各学科领域和交叉学科的会议录和期刊文献进行快捷报道;<http://www.ieee.org> 是检索会议文献的热门站点;<http://lib-www.ucr.edu> 拥有 5000 多条索引,收录了 1.1 万多种学术期刊等。

5) 专业学术 FTP。很多高校均有自己的专业 FTP,这些 FTP 往往具有大量的学术性资源需要注意搜集、利用。这其中用得最多的是有关共享软件免费的发布,对学校的计算机教学十分有益,应注意搜集。

4.3 采集途径与方式

1) 订购。为了做好订购工作,要与出版发行部门经常取得联系,主动介绍本单位的业务情况及对数字信息资源的各种需求,以便能得到最新、最全的信息资源。

2) 交换。这是信息资源交流的一种形式,也是资源收集的重要途径。交换可以获得不公开发行或不能通过贸易途径获得的非卖品,便于补缺。尤其是在

网络资源交换上,大家都已形成共识,利用 BT、eDonkey 等 P2P 方式交换资料就是一种非常好的信息搜集方式。

3) 现场搜集。通过参观访问或参加国内外各种学术会议、经验交流会、各种讲座等方式,进行文献的现场搜集也是一种有效的办法。比如利用数码摄像机进行现场拍摄、利用录音笔进行现场录音可以直接取得第一手资料。而一般来说,高校学术讲座数量相当多,我校仅去年就举办了上百场次的学术讲座。利用这种方式获取的资源,不仅用时少、可靠性强、实用率高,而且针对性强,便于学校大面积传播学术信息的需要。

4) 索取。有些资源免费赠送,例如国内外各种产品展览会上产品说明书等,可以向主办者索取;需要什么资源,还可以在网络上发帖子,会得到帮助。本单位教师教学成果、研究成果等,可以向教师索取。

5) 委托收集。对于一些学科前沿的资源,可以委托本学科的专家、学者收集。

6) 转换。将其他形式的数字资源转换成系统需要的格式统一的数字资源。比如很多单位为了网络发布的需要,将 VCD、DVD、MPEG-4 格式的视频数字文献统一转化为 RMVB 格式,将 CD、WAV、ASF、RA 等格式的音频数字文献统一转换为 MP3 格式……,从而实现网络的快速传输。

4.4 数字资源的选择方法

1) 根据搜索引擎对文献的评介来选择。作为上网使用的最主要的一项工具,搜索引擎对文献的评价也有举足轻重的影响,在一定程度上可能直接决定用户对某一文献的取舍。就目前来说,搜索引擎对文献的评介主要体现在分类目录和搜索结果两个方面。

2) 根据新闻组对文献的评介来选择。新闻组则是互联网上一种高效的交流方式,有关图书评介的新闻组也是俯拾皆是。比如万千新闻组有文化艺术·文学、艺术·文学·诗词、“教育”组等。大洋网有个人·freefish、文学天地、个人·bravebull、优秀主页推荐、个人·maylan、金庸客栈、个人·Y.K. 网络文学。这些组中均含有大量的文献评介。

3) 根据网络书评对文献的评介来选择。“书评”在网络中也得到了迅速的发展,如黄金书屋的“书评麻辣烫”、紫杉书房的“推荐与书评”、金陵图书馆的“快语书评”、象牙塔:国史探微:书评书讯、百博书评、黄金书屋的书评总汇(量大)、China-pub.com 书评在线等。一般来说,这类站点的书评的总体水平较

高,特别是一些高水平的书评往往被广为转贴,也成为学术研究的重要材料。

4) 根据邮件列表对文献的评介来选择。作为一种重要的信息传播工具,邮件列表在文献评论方面也发挥了重要作用。比如黄金书屋每周更新邮件中有本期推荐、每日更新、网文精选、人气排行等栏目,在这些栏目中均有各种文献评介。再比如大洋网的邮件列表——每周靓书中也有文献评价。这些评价的影响力比较大,往往直接影响阅读倾向,也即具有导读效果。

5) 通过传统书刊的评价来选择。传统书刊中往往含有大量的数字资源的评论以及相关的广告,这些资料也可以成为我们衡量数字资源质量的一个重要依据。

5 数字信息采集工作中需要注意处理的几个关系

5.1 处理好重点与一般的关系

数字信息资源的采集必须要有重点,根据本校教学、科研工作需要的主要学科或技术领域选择重点,确定重点要考虑到当前和长远需要。重点资源的品种的系统、完整,以保证满足工作的需要。所谓照顾一般,就是在保证重点资源收藏的前提下,对于与本学科或技术领域相关的或用途广泛的一般资源,也要适当收藏,但在范围和数量上要适当控制。

5.2 处理好数字资源与非数字资源的关系

信息媒介正在相互靠拢、相互融合。在某些情况下它们之间的界限正在逐步消失。例如同一信息产品可寓于书本型、缩微品、光盘、录像带、通信传输等多种介质之中。可由出版社、图书馆、通讯社、数据库服务、联机检索等途径提供可通过电视、广播、电子邮件、卫星通信等传播。信息媒介日趋多样化和相互融合构成了现代社会电子信息资源的另一个令人瞩目的发展趋势。

5.3 处理好当前与长远的关系

数字信息资源的采集不仅考虑当前的需要,还要考虑到发展的需要,收集人员应密切注意各门学科的水平、动向和发展趋势,随时留意最新动态,尽可能照顾本单位的长远发展规划的需要,只有这样,才能使资源收集工作走在教学、科研工作的前面。

5.4 处理好与其他业务部门的关系

数字信息资源的采集是电子图书馆的一个重要组成部分,与数字信息资源的加工部门、出版发行部门、财务部门、读者服务部门等有着广泛的联系。因此,与这些部门之间的协调、合作是非常必要的。